A LOOK AT A DAY OF DATA FROM THE TOPEX/POSEIDON GPS RECEIVER

J. Rodney Jee, California Institute of Technology
Jet Propulsion Laboratory, 4800 Oak Grove Drive, MS 301-125J, Pasadena, CA 91109

KEY WORDS: robust regression spline, graphics

INTRODUCTION

Launched on August 10, 1992, the Topex/Poseidon satellite is a joint project of NASA and the French space agency CNES. The purpose of this Earth orbiting altimetric satellite (for brevity, referred to as Topex) is to enable monitoring of the ocean topography to higher levels of accuracy than previously achieved. Essential to obtaining this greater level of accuracy is the use of improved navigational systems which can yield position estimates for Topex good to a few centimeters. One of the navigational systems utilized by Topex is the Global Positioning System (GPS) a network of navigational beacons that has become widely used in many areas beside spacecraft navigation. Topex is one of the first satellites to utilize a high precision dual frequency GPS receiver and the status of the use of the GPS for Topex is that of an experiment rather than as the primary operational system for navigation. In addition to the on-board receiver, the navigation experiment requires a network of ground based GPS receivers to collect data simultaneously with Topex.

The focus of this talk is on the data pre-processing (specifically, the data editing) that must be performed on the GPS data from Topex before it is passed to the orbit determination part of the software. The editing is done by a program that is part of the GPS Data Processing Facility (GDPF)—a set of software designed and used by the Jet Propulsion Laboratory for this navigation experiment. It should come as no surprise that some form of data editing must be performed initially. The main goal of the data editor is to relieve the orbit analyst of further need for data editing- especially in the orbit determination stage where problems arising from faulty data are more difficult and computationally expensive to diagnose. Numerous computer programs for editing GPS data are in use within the GPS user community (see the papers in Bock and Leppard 1990). The data editor of the GDPF has incorporated some of the ideas of these programs while adding in some additional and improved procedures. As an integral part of its processing the GDPF data editor maintains a statistical summary of all processing that is performed. In addition it automatically generates a fairly complete set of diagnostic graphics in postscript form which may be printed or viewed on a computer monitor. A sample of this graphical output will be presented here along with some highlights of the workings of the data editor. A goal of this talk is to show the role that good graphics and statistics has in facilitating the processing and understanding of large amounts of data.

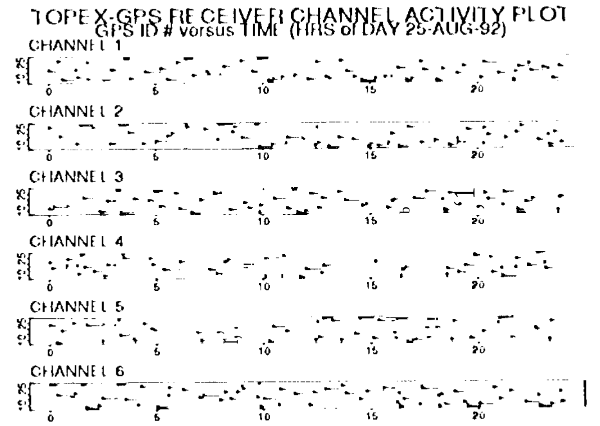

FIGURE 1: FIRST FULL DAY OF DATA FROM THE TOPEX-GPS RECEIVER

A BRIEF DESCRIPTION OF GPS DATA

The Topex Design Receiver (DR) has six channels to allow it to simultaneously track 6 GPS satellites (see Figure 1). These 6 satellites are optimally selected at regular intervals by the receiver from a constellation of over 24 (not all of which are in view at any instant). Each of the 6 channels records a multivariate time series which I now describe briefly. For each GPS beacon the pseudorange data type is available on carriers at 2 frequencies (1575.42GHz and 1227.6GHz, respectively, or about 19cm and 24cm) at the rate of 1 point per 10 seconds each. This data type is most naturally thought of as a measurement giving the time it takes the signal to travel from the GPS transmitter to the receiver. Its units are time, but it is common to convert pseudorange to units of length (by multiplying by the speed of light) which is done in the GDPF. These data have decimeter level precision and will be referred to as P1 and P2. For each GPS the phase data type is available at the same two frequencies as pseudorange but at the higher rate of once per second. This data type can be thought of as accumulated counts of cycles of the received signal. Multiplication by the appropriate wavelength converts these to units of length. These data have millimeter level precision anti will be designated 1.1 and 1.2. The set of 1'1, 1'?, 1.1, and 1.2 data recorded by uninterrupted, continuous tracking of a particular GPS is referred to as a "pass" of data. Typically, the DR collects over 300 passes of data a day. Since the DR continuously collects phase data every second simple arithmetic shows that over 1 million points per day must be processed. The differential span of both the pseudorange and phase data during a pass can be about 8000 km. The need to retain better than millimeter precision in the data thus requires high precision arithmetic.

STATISTICAL, METHODS & SOFTWARE

All of the data are fit (per time series, per pass) with robust regression splines. The purpose in performing the splint fits is to produce residuals as one means to spot faulty data-outliers or other specific problems known to occur for the data. The GDPF provides over 20 user specifiable parameters for controlling how the splint fits arc done. Among these are splint parameters for selling (for each data type) the constant knot spacing, the constant knot repetition to control the degree of derivative. continuity, and the constant degree of the polynomial splints. 'J'here arc also parameters for controlling the iteratively reweighted least squares algorithm used to solve for the splines minimizing the Huber weighting criterion (Huber 198 1). Bivariate discriminant analysis is performed with robust covariance m atrices calculated with Median Absolu te Deviation (MAD) based variances and covariances using the identity

$$COV(X,Y) = (VAR(X+Y) - VAR(X-Y)) / 4.$$

These algorithms arc implemented in Fortran in the GDPF which is run in a batch mode.

Capability for interactive analysis of the automated editor's results is built cm top of the S-1'1,11S software (Statistical Sciences, Inc. 1992). This version of the S language (Becker, Chambers, and Wilks 1988) provides a Fortran interface for running subroutines of the data editor with convenient access to all arguments and results, graphical capabilities (X-windows and postscript), and a flexible programming environment used to perform special analyses and prototyping of algorithms for Fortran implementation.

PHASE DATA COMBINATIONS

Figure 2 shows plots of one pass of phase data in the linear combinations that arc usually formed by GPS data analysts. The top plot shows the linear combination of phase data often designated LC (= 1.1 + 1.54(1 ,1-1,2)). This is the only combination of the phase data which is used by the orbit determination part of the Gill'1' as the duai frequencies arc for ionospheric calibration purposes (to be made more clear soon). Since the differential span of the LC is typically thousands of kilometers while phase problems arc often sub-meter in level, plots of LC seldom reveal any problems discernible by the unaided eye. The bottom plot of Figure 2 shows the ionospheric combination 1.1-1,2. This linear combination is dominated by ionospheric effects and is of a scale as to allow one to sometimes scc directly the level of the noise in the data. 'I'here arc some noticeable "discontinuities" in the 1.1-1,2 plot of Figure 2. They arc manifestations of what arc known as cycle slips- the main problem the GPS data editors must contend with. It has turned out that the Topex-GPS Receiver rarely has cycle slips, so the data for Figure 2 is not chosen as a representative sample but rather as one of the few that does have the cycle slip problem.

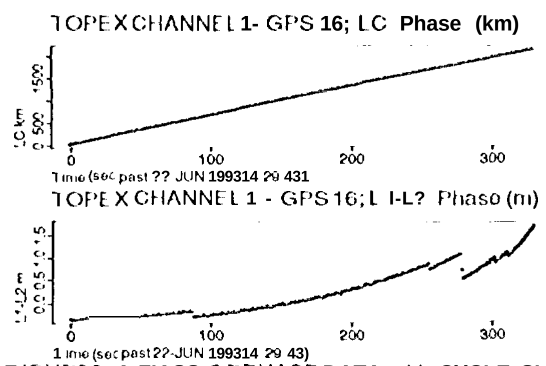

FIGURE 2: A PASS OF PHASE DATA with CYCLE SLIPS

THE CYCLE SLIP PROBLEM

As illustrated in Figure 2 the cycle stip problem results in discontinuities in the phase data. The sizes of the discontinuities arc multiples of a half-integer when the data arc in units of cycles in the original 1.1 and 1.2 (not in the LC and 1,1-1.2 combinations).

The data editing software tackles the cycle slip problem by forming the divided differences of the data in time thereby converting the discontinuities into outliers. Then robust regression splines arc. fit to the rate(LC) and rate(L1-L2) to produce residuals by which cycle slips are identified. Scc Figure 3 for time series plots of the rates of the phase. The bottom plot of Figure 3 overlays the splint fit on the. data Values.

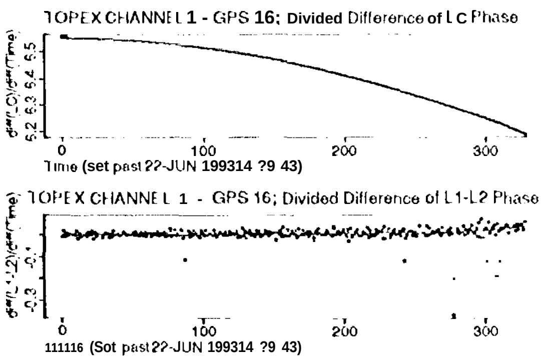

FIGURE 3: RATE of PHASE DATA and SPLINE FIT

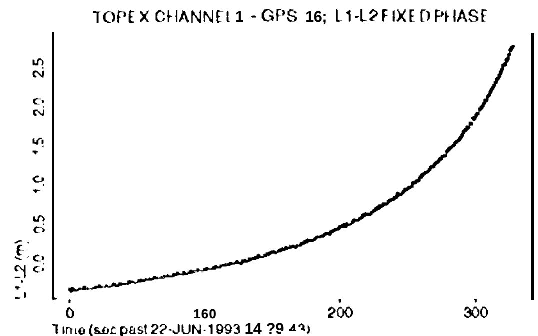

FIGURE 5: PHASE DATA CORRECTED FOR SLIPS

It is highly desirable to fix the cycle slips if possible. The data editing software fixes cycle slips by using two-dimensional discriminant analysis to resolve how many half-integers were jumped. Figure 4 shows a plot of the residuals from splint fits to the data of Figure 3 in a scatter-plot along with contours of equal probability centered at half-integer locations. Note that the residuals in rate(LC) and rate(L1-L2) have been transformed to units of cycles in L1 and L2. As usual in discriminant analysis the contours are based upon Gaussian assumptions but using robust covariance estimation as indicated earlier. The cycle slip identification algorithm proceeds by first determining the cycle slip center to which a point is closest. Then it decks that a point is within some user specified acceptance region to bound the probability of misclassification. If a point cannot be classified with very low probability of error (as specified by the user), then the cycle slip is simply marked as the beginning of a new pass. in this case all the cycle slips were properly identified and the fixed L1-L2 combination is shown in Figure 5.

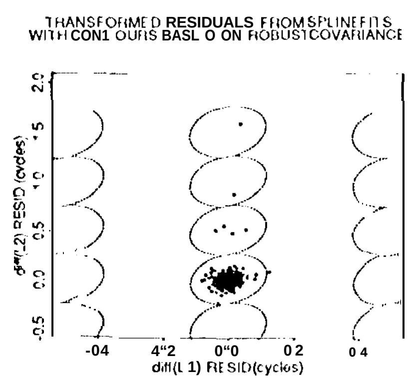

FIGURE 4: CYCLE SLIP DISCRIMINATION

STATISTICS FROM THE EDITOR

The data editor calculates and saves over 30 statistics for every pass of data it successfully processes. These statistics provide a means of monitoring the editor's performance and of monitoring the receiver's activities; unusual statistics could signal that processing is invalid or that some unanticipated problem slipped through the editor. An examination of these statistics for each of the 300 plus passes per day is the main way the editor's performance is validated before the edited data are passed on to additional processing. Of course, graphical representation of these statistics greatly aids in understanding. For each day of data processing the editor automatically produces over 10 postscript files of graphics (most with multiple plots) of these statistics. Figure 1 is a simplified version of one of these daily graphical summaries.

As another example, Figure 6 shows a scatterplot of the MADs of all the passes of residuals from the spline fits to LC and L1-L2 phase combinations for 31 May 1993. A point with extreme statistics appears in this plot and signals that the pass should be inspected for some type of failure in the editing. The L1-L2 combination for that pass is plotted in Figure 7. For a GPS data analyst this plot reveals some obvious problems with the data that had not been anticipated when designing the editor. The appropriate action for this pass is to delete all of the data and determine whether it is the receiver or the transmitter which is at fault. As it turned out, GPS 31 was experiencing hardware problems on that day.